VISIT…

LANZAROTE Caliente.com

НЕКОТОРЫЕ ОЦЕНКИ ЭФФЕКТИВНОСТИ ПАРАЛЛЕЛЬНЫХ АЛГОРИТМОВ РЕШЕНИЯ СЛАУ НА ПОДПРОСТРАНСТВАХ КРЫЛОВА*

М. Ю. Баландин, Э. П. Шурина
Новосибирский государственный технический университет, Россия
e-mail: shurina@fpm.nstu.nsk.su

The ABR1ORT method for solving densed ill-conditioned systems of linear equations (SLEs) and the GMRES method for solving sparsed non-symmetric SLEs are discussed; both them belong to Krylov Sequence Methods. The analysis of numerical complexity and memory requirements is performed for these methods; parallel algorythms for multiprocessor computers are briefly discussed. For these algorythms, asymptotic estimations of parallel speedup are shown.

1. Введение

Решение ряда задач математической физики (задачи гидрогазодинамики, расчета электромагнитных полей, уравнения Максвелла, Навье — Стокса и др.) методами конечных элементов (FEM) и конечных объемов (FVM) — особенно на неструктурированных сетках — приводит к системам линейных алгебраических уравнений с разреженными матрицами большой размерности, которые, как правило, являются несимметричными [10]. Частным случаем несимметричности является симметричный портрет с асимметрией коэффициентов, т. е. *одинаковое* расположение *различных* коэффициентов матрицы по ее верхнему и нижнему треугольникам [5]. В задачах гидрогазодинамики исключение давления нередко приводит к необходимости обращения близких к вырожденным (и, как следствие, плохо обусловленных) СЛАУ с плотной матрицей; плохая обусловленность может также вызываться погрешностями задания коэффициентов матрицы, связанными с неточностями измерений и/или аппроксимацией. К СЛАУ с плотной плохо обусловленной матрицей приводят также задачи, формализуемые в виде интегральных уравнений 1-го и 2-го рода [2].

Эффективным средством решения задач большой размерности являются многопроцессорные вычислительные системы, однако соответствующая реализация классических методов требует их специальной адаптации и проведения связанных с этим исследований [5, 10, 6].

В настоящей работе рассматриваются метод обобщенных минимальных невязок GMRES, эффективный для решения разреженных СЛАУ с несимметричной матрицей

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, грант №96–01–01557.

© М. Ю. Баландин, Э. П. Шурина, 1998.

[4, 6, 8], и метод ABR1ORT из класса методов, предложенных А. А. Абрамовым для решения плотных плохо обусловленных СЛАУ [1, 3, 7]. Оба метода основываются на проектировании искомого решения на образуемое системой подпространство Крылова.

Целью работы является создание параллельных алгоритмов методов и теоретическая оценка их эффективности с точки зрения увеличения скорости решения по сравнению с последовательной реализацией; алгебраические вопросы улучшения сходимости в рамках данного исследования не рассматриваются. Для ABR1ORT и GMRES приводятся оценки требуемого объема памяти, вычислительной сложности и асимптотического ускорения счета для параллельных реализаций по SPMD-схеме [5].

2. Методы

Будем предполагать, что решаемая СЛАУ имеет вид $\mathbf{A}x = b$, где $\mathbf{A}$ — матрица размерности $n \times n$, а x и b — векторы размерности $n \times 1$. Пусть a_i — i-я строка матрицы $\mathbf{A}$ (вектор $1 \times n$), а b_i — i-й компонент вектора b. Предполагается также, что процедура предобусловливания проводится отдельно, и если она необходима, то метод применяется к уже измененной СЛАУ (так называемое "внешнее" предобусловливание), после чего найденное решение пересчитывается с учетом ранее выполненного предобусловливания. (Может использоваться, например, процедура вида $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}y = \mathbf{D}^{-1/2}b$, где D — диагональная матрица с главной диагональю, взятой из матрицы $\mathbf{A}$.) Говоря о параллельных реализациях методов, будем обозначать доступное число процессоров как p.

Предложенный А. А. Абрамовым класс методов [1] основан на проектировании искомого решения на линейное пространство, образованное строками матрицы, и последующем уменьшении размерности этого пространства; сумма найденных проекций, таким образом, даст решение СЛАУ. Указанные методы требуют пересчета матрицы и потому пригодны для решения лишь плотных СЛАУ, однако, как было показано, позволяют достичь весьма высокой точности.

В общем виде данный класс может быть записан как уточнение решения по формулам:

$$x_0 = 0, \quad x_{i+1} = x_i + \phi_{i+1} d_{i+1}, \quad \phi_i = \frac{\psi_i}{(d_i\, d_i)},$$

$$a_j^{i+1} = a_j^i - \omega_j (d_i)^T, \quad b_j^{i+1} = b_j^i - \omega_j \psi_i,$$

$$\omega_j = \frac{(a_j^i\, d_i)}{(d_i\, d_i)},$$

$$d_i = (\mathbf{A}^i)^T g_i, \quad \psi_i = (b^i)^T g_i, \tag{1}$$

где вектор g_i является параметром, выбор которого и определяет конкретный метод из класса.

Подробное описание предложенного А. А. Абрамовым класса методов и его геометрическая интерпретация приведены в работе [7].

Метод ABR1ORT получается при $g_i = e_i = (0, 0, \ldots, 1_i, 0, \ldots, 0)^T$. Формулы (1) при этом могут быть упрощены, так как $d_i = (\mathbf{A}^i)^T e_i = a_i^i$ и $\psi_i = (b^i)^T e^i = b_i^i$. Кроме того, в таком варианте метод становится чувствительным к критерию останова [3], и для повышения точности целесообразно применение специального условия, определяющего допустимость

включения очередного вектора в базис подпространства Крылова:

$$\frac{|(b^k_{j(k+1)}a^k_{j(k)} - b^k_{j(k)}a^k_{j(k+1)} \quad a^k_{j(k)})|}{\|a^k_{j(k)}\|^2} < \varepsilon,$$

где $j(s)$ —номер базовой строки на s-й итерации.

Метод GMRES [4, 6, 9] был предложен для решения несимметричных СЛАУ и основан на построении базиса в соответствующем системе подпространстве Крылова $K^j(A, r_0)$. Затем решение уточняется некоторой добавкой, представленной в виде разложения по этому базису.

В несимметричном случае основная сложность заключается в том, что для построения ортогонального базиса не могут быть использованы рекуррентные соотношения с малой глубиной зависимости типа тех, что применяются в методе сопряженных градиентов. Поэтому в GMRES ортонормальный базис $\{v_j\}$ строится на основе следующего соотношения, выражающего полную зависимость всех векторов:

$$h_{i+1,i}v_{i+1} = \mathbf{A}v_i - \sum_{j=1}^{i} h_{j,i}v_j, \quad v_1 = \frac{1}{\|r_0\|}r_0. \tag{2}$$

В матрично-векторных обозначениях соотношение (2) может быть записано как

$$\mathbf{AV}_i = \mathbf{V}_i\mathbf{H}_i + h_{i+1,i}v_{i+1}e_i^T = \mathbf{V}_{i+1}\overline{\mathbf{H}}_i, \tag{3}$$

где $\mathbf{V}_i = [v_1|v_2|...|v_i]$, а $\overline{\mathbf{H}}_i$ —матрица размерности $(i+1) \times i$ в верхней форме Хессенберга.

Особенностью GMRES является то, что с целью экономии памяти выбирается некоторая размерность подпространства Крылова m (причем $m \ll n$), и после того как m базисных векторов построены решение уточняется. Построение очередного вектора v_i обычно называют GMRES-итерацией, а m таких итераций с уточнением решения — GMRES-циклом. Один цикл требует $m + 1$ матрично-векторное умножение.

Система (3) является плотной и имеет малые размеры; кроме того, она переопределена, а потому должна решаться в смысле наименьших квадратов.

Следует также отметить, что при $m = n$ GMRES фактически становится методом абрамовского класса. Подробное описание GMRES приведено в [6, 8, 9].

3. Анализ методов, параллельные алгоритмы и асимптотические оценки

Схема метода ABR1ORT характерна тем, что объем вычислительной работы от итерации к итерации уменьшается, так как в качестве векторов d_i последовательно выступают строки матрицы $\mathbf{A}$, которые в дальнейших расчетах уже не участвуют. Основная тяжесть вычислительной работы приходится на вычисления скалярных произведений строк матрицы и их линейное комбинирование. В общей сложности на i-й итерации вычисляется $n - i + 1$ скалярное произведение (i-я строка на себя и на каждую из $n - i$ нижележащих) и производится $n - i$ операций линейного комбинирования строк.

Несложные подсчеты показывают, что на i-й итерации с учетом проверки значимости строк выполняется в общей сложности $4n^2 - 4(n-1)i + 7n + 1$ арифметическое действие.

Поскольку

$$\sum_{i=1}^{n}(4n^2 - 4(n-1)i + 7n + 1) = 2n^3 + 7n^2 + 3n = N^{\mathrm{A}}(n),$$

то можно говорить, что метод имеет сложность $O(n^3)$. Действительно, его вычислительная схема обработки строк матрицы схожа со схемой метода Гаусса, который имеет такую же сложность.

Интерес представляет также исследование метода с точки зрения требуемой им машинной памяти. При программной реализации метод требует хранения матрицы СЛАУ (использование разреженных форматов нецелесообразно, так как матрица постоянно пересчитывается), т. е. n^2 вещественных чисел; векторов решения и правой части, т. е. $2n$ вещественных чисел; значений скалярных произведений фиксированной строки матрицы на саму себя и на все остальные строки, т. е. n вещественных чисел; вектора флагов значимости строк матрицы, т. е. n логических величин. Таким образом, всего требуется хранение $n^2 + 3n$ вещественных и n логических величин (объем памяти под счетчики циклов и временные переменные скалярного типа можно считать пренебрежимо малым). Если вещественное число требует 4 байт, а логическая величина 1 байт, то для решения СЛАУ размерности $n \times n$ метод ABR1ORT требует $4n^2 + 13n$ байт; при использовании вещественных чисел двойной точности требования возрастают до $8n^2 + 25n$ байт. Следовательно, требования метода ABR1ORT к объему доступной памяти составляют $O(n^2)$.

Оценить вычислительную сложность метода GMRES несколько труднее, так как он является итерационным, а не прямым, причем количество итераций определяется размерностью GMRES-цикла. Можно оценить лишь сложность одного цикла как функцию его размерности. Если предполагать, что матрица решаемой СЛАУ является разреженной, то необходимо еще одно обозначение — количество внедиагональных ненулевых элементов в матрице. Пусть несимметричная матрица $\mathbf{A}$ имеет симметричный портрет и число ненулевых элементов в нижнем (верхнем) треугольнике равно k.

Если решение малой переопределенной СЛАУ производится с помощью QR-факторизации с применением вращений Гивенса [8–10], то подсчет показывает, что общее количество арифметических действий, выполняемых на одном GMRES-цикле,

$$N^{\mathrm{G}}(n,m) = 5n + 9m + 6k + 8nm + 2nm^2 + 2m^2 + 2mk + 1.$$

С учетом того, что $m \ll n$ (возможности современных вычислительных систем обычно ограничивают выбор размерности подпространства Крылова величиной $m \cong 0.01n$) и $k = O(n)$ (обычно $k \leq 100n$), можно говорить, что один GMRES-цикл имеет сложность $O(nm^2)$. Проблема оптимального автоматического выбора размерности подпространства Крылова m в настоящее время не решена, однако число GMRES-циклов теоретически не превосходит $\lceil n/m \rceil$.

При выборе $m = n$ для решения СЛАУ становится достаточно только одного GMRES-цикла, и метод приобретает сложность $O(n \cdot n^2) = O(n^3)$. Действительно, данная величина совпадает со сложностью метода ABR1ORT.

При анализе требований GMRES к объему доступной памяти также будем предполагать, что матрица СЛАУ несимметрична, но имеет симметричный портрет (это позволяет снизить расходы памяти на ее хранение за счет того, что хранится портрет только одного — верхнего либо нижнего — треугольника матрицы).

Вычислительная схема требует работы со следующими объектами: разреженной матрицей исходной СЛАУ — $n + 2k$ вещественных и $n + k + 1$ целых (причем обычно двойной

длины!) чисел; векторами правой части, искомого решения и невязки — $3n$ вещественных чисел; векторами базиса подпространства Крылова — mn вещественных чисел; плотной матрицей малой переопределенной СЛАУ — $m(m+1)$ вещественных чисел; коэффициентами вращений Гивенса — $2m$ вещественных чисел; вектором решения малой СЛАУ (как с коэффициентами линейного комбинирования векторов базиса подпространства Крылова) — m вещественных чисел.

Полагая, как и прежде, объем памяти для счетчиков циклов и временных переменных скалярного типа пренебрежимо малым, можно получить, что всего требуется хранить $n+k+1$ целых чисел двойной длины и $4n+2k+mn+(m+1)^2+3m+1$ — вещественных. При использовании обычной точности это требует $20n+12k+4mn+4(m+1)^2+12m+8$ байт, двойной — $36n+20k+8mn+8(m+1)^2+24m+16$ байт. С учетом того, что $1 < m \ll n$ и $k \ll n$, можно говорить, что требования метода GMRES к объему доступной памяти составляют $O(mn)$.

Как было показано ранее, основные вычислительные затраты в ABR1ORT определяются операциями нахождения скалярных произведений и линейного комбинирования векторов. Можно подсчитать, что эти действия требуют $2n^3+\dfrac{3}{2}n^2-\dfrac{1}{2}n$ арифметических операций. Доля соответствующих вычислительных затрат в схеме метода (требующей $N^{\mathrm{A}}(n)$ операций), таким образом, равна

$$\frac{2n^3+\dfrac{3}{2}n^2-\dfrac{1}{2}n}{2n^3+7n^2+3n},$$

т. е. при больших размерностях СЛАУ близка к 1. При этом строки матрицы СЛАУ, начиная с первой, последовательно исключаются из процесса решения. В такой ситуации наиболее эффективной является параллелизация, основанная на столбцовой декомпозиции матрицы СЛАУ при реализации вычислений на топологии межпроцессорных соединений "каждый с каждым". При этом работу с вектором правой части целесообразно дублировать на каждом из процессоров. Анализ показывает, что при использовании такой схемы на p процессорах число арифметических действий, выполняемых одним процессором, сокращается до величины

$$N_p^{\mathrm{A}}(n) = \frac{n(3n^2+2n+np+p+1)}{p}.$$

Таким образом, асимптотическое ускорение для ABR1ORT равно

$$\lim_{n\to\infty}\frac{N^{\mathrm{A}}(n)}{N_p^{\mathrm{A}}(n)} = \frac{2}{3}p. \tag{4}$$

При этом следует иметь в виду, что в реальных задачах, когда между строками матрицы существует близкая к линейной зависимость (т. е. матрица вырождена или близка к вырожденной), ускорение может заметно отличаться от полученной линейной зависимости за счет возрастания объема межпроцессорных коммуникаций, не связанных с вычислениями (так как в передаваемых наборах попарных скалярных произведений строк СЛАУ некоторые значения реально не будут использованы; при этом передача их несколькими коммуникационными пакетами не снимает проблемы, а наоборот, лишь увеличивает время передачи [5]).

Численные эксперименты (см. раздел 4) на матрицах Гильберта (наихудший случай плохо обусловленных матриц) показали результаты, довольно хорошо согласующиеся с (4).

Вычислительная схема GMRES значительно более сложна. Как можно видеть из сказанного ранее, в ней присутствуют два интенсивных в вычислительном плане этапа: умножение матрицы СЛАУ на вектор (для нахождения невязок) и нахождение скалярных произведений векторов (при построении ортонормального базиса подпространства Крылова). Кроме того, именно на хранение векторов базиса $\{v_j\}$ в GMRES приходится бóльшая часть затрат памяти (mn вещественных чисел, что часто оказывается больше объема памяти под разреженную матрицу СЛАУ). При этом в GMRES-схеме присутствует достаточно самостоятельная подзадача — решение переопределенной СЛАУ малой размерности в смысле наименьших квадратов, которая должна рассматриваться отдельно.

Перечисленные факторы приводят к следующим выводам о параллельной реализации алгоритма. Для нахождения матрично-векторных произведений при вычислении невязок матрица СЛАУ должна быть подвергнута декомпозиции на строчные блоки, размеры которых определяются требованиями равномерности загрузки. Соответствующим образом подвергается декомпозиции и вектор правой части. Для наиболее эффективного вычисления скалярных произведений и линейных комбинаций векторов базиса каждый из векторов v_j должен быть разбит на блоки равного размера по числу процессоров; каждый из процессоров при этом хранит соответствующие блоки всех векторов. Обработка вектора решения дублируется на всех процессорах.

Используемая топология в значительной степени определяется тем, каким способом решается малая переопределенная СЛАУ. Если данная подзадача решается последовательно (что вполне возможно из-за ее малых размеров), то наиболее подходящей оказывается топология “звезда”.

Пусть малая СЛАУ решается последовательно с применением метода вращений Гивенса. Тогда анализ для топологии “звезда” показывает, что за один цикл вычислительная нагрузка центрального (наиболее занятого) процессора “звезды” составляет

$$N_p^{\mathrm{G}} = \frac{4n + 8k + 1 + 6mn + 2mk + 2m^2n + 2m + 2m^2p + 2mp + 2mnp}{p}$$

арифметических операций.

Таким образом, асимптотическое ускорение, равное

$$\lim_{n\to\infty} \frac{N^{\mathrm{G}}(n,m)}{N_p^{\mathrm{G}}(n,m)} = \frac{1}{2} \cdot \frac{(5 + 8m + 2m^2)p}{2 + 3m + m^2 + mp}, \tag{5}$$

оказывается функцией не только числа процессоров, но и размерности подпространства Крылова m ($m \ll n$). Так, можно видеть, что при больших m ($m \to n$) асимптотическое ускорение составит

$$\lim_{n\to\infty} \left(\lim_{m\to n} \frac{N^{\mathrm{G}}(n,m)}{N_p^{\mathrm{G}}(n,m)} \right) = p.$$

Формула (5) показывает, что выбор глубины пространства Крылова является весьма важной проблемой. Теоретически выбор большего значения должен приводить к более быстрому достижению решения, однако это справедливо лишь для вычислений в точной арифметике. Как показали численные эксперименты, при чрезмерном увеличении m возможно даже замедление расчетов, так как метод становится чрезвычайно чувствительным

к вспомогательному критерию точности, который на заключительных циклах перестает распознавать "избыточные" векторы в базисе подпространства Крылова.

4. Некоторые результаты численных экспериментов

В табл. 1 и 2 показаны значения ускорения при параллельном счете методами ABR1ORT и GMRES в сравнении с теоретическими оценками. Эксперименты проводились на вычислительной системе PARSYTEC в среде Power Xplorer (см. также [3]). Обозначения: t_p — время решения на процессорах; $s_p^{\text{теор}}$ — значение теоретической оценки ускорения (см. раздел 3); $s_p^{\text{прак}} = t_1/t_p$ — практически достигнутое ускорение. Погрешность измерения времени при расчетах составляла ± 1с.

Метод ABR1ORT исследовался на симметричных матрицах Гильберта, GMRES — на матрице размерности 8112, полученной из задачи расчета электромагнитного поля методом конечных элементов (подробности см. в [8]). При этом расчеты производились без предобусловливания и с диагональным предобусловливанием вида $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}y = \mathbf{D}^{-1/2}b$.

Т а б л и ц а 1. Метод ABR1ORT.

n	p	t_p	$s_p^{\text{теор}}$	$s_p^{\text{прак}}$
700	1	$6''$	1.0	1.0
	2	$6''$	1.3	1.0
	3	$5''$	2.0	1.2
	4	$3''$	2.7	2.0
800	1	$7''$	1.0	1.0
	2	$6''$	1.3	1.2
	3	$5''$	2.0	1.4
	4	$4''$	2.7	1.8
900	1	$10''$	1.0	1.0
	2	$8''$	1.3	1.3
	3	$6''$	2.0	1.7
	4	$5''$	2.7	2.0

Т а б л и ц а 2. Метод GMRES.

m	p	t_p		$s_p^{\text{теор}}$	$s_p^{\text{прак}}$	
		б/предоб.	предоб.		б/предоб.	предоб.
10	1	$11'36''$	$10''$	1.00	1.00	1.0
	2	$8'27''$	$7''$	1.88	1.37	1.4
	4	$5'6''$	$4''$	3.31	2.28	2.5
20	1	$7'8''$	$9''$	1.00	1.00	1.0
	2	$4'59''$	$6''$	1.92	1.43	1.3
	4	$3'22''$	$4''$	3.56	2.12	2.3
30	1	$5'28''$	$8''$	1.00	1.00	1.0
	2	$4'1''$	$6''$	1.94	1.36	1.3
	4	$2'37''$	$3''$	3.68	2.09	2.7

5. Заключение

В процессе исследований был детально проанализирован и эффективно реализован метод ABR1ORT, пригодный для решения плотных плохо обусловленных СЛАУ с высокой точностью (в том числе недо- и переопределенных систем), а также реализован и исследован метод GMRES; в связи со сложностью схемы GMRES представляются целесообразными эксперименты с реализацией метода на системах с разделяемой памятью [5].

Численные эксперименты показали достаточно хорошую (с учетом сравнительно низкой скорости передачи данных по межпроцессорным соединениям у вычислительной системы PARSYTEC [6]) согласованность с теоретическими оценками (4), (5)

Дальнейшие исследования в данном направлении связаны в первую очередь с применением ABR1ORT в схеме GMRES и проблемой автоматического выбора размерности подпространства Крылова, а также с использованием этих методов для решения нелинейных систем.

Список литературы

[1] Абрамов А. А. Об одном методе решения плохо обусловленных систем линейных алгебраических уравнений. *Журн. вычислит. матем. и матем. физики*, **31**, №4, 1991, 483–492.

[2] Абрамов А. А. О свойствах метода Крейга при решении линейных некорректных задач. *Там же*, **35**, №1, 1995, 144–150.

[3] Баландин М. Ю., Шурина Э. П., Чернышев О. В. Анализ параллельных алгоритмов решения плохо обусловленных СЛАУ. *Тр. III междунар. науч.-техн. конф. "Актуальные проблемы электронного приборостроения" (АПЭП-96)*, Т. 6, Ч. 1, Новосибирск, 1996, 74–77.

[4] Ильин В. П. *Методы неполной факторизации для решения алгебраических систем.* Физматлит, М., 1995.

[5] Ортега Дж. *Введение в параллельные и векторные методы решения линейных систем.* Пер. с англ. Мир, М., 1991.

[6] *Aspects of Computational Science.* Stichting Nationale Computer Faciliteiten, The Netherlands, 1995.

[7] Balandin M. Yu. et al. On parallelization of new algorithm for solving systems of linear equations. In *"Joint Bulletin of the Novosibirsk Computing Center and the Institute of Informatics Systems"*, Ser. "Computer Science", 1996, 17–26.

[8] Balandin M., Chernyshev O., Shurina E. Analysis of methods for solving large-scale non-symmetric linear systems with sparsed matrices. In *"Parallel Computing Technologies, 4th Int. Conf. (PaCT-97) Proc."* Springer-Verlag, "Lecture Notes in Computer Science" series, **1277**, 1997, 336–343.

[9] Kadioglu M., Mudrick S. On the implementation of the GMRES(m) method to elliptic equations in meteorology. *J. of Comput. Phys.*, **102**, 1992, 348–359.

[10] Natarajan R. Finite element applications on a shared-memory multiprocessor: algorithms and experimental results. *Ibid.*, **94**, 1991, 352–381.

Поступила в редакцию 18 ноября 1997 г.